

# Information-theoretic measurements of coupling between structure and dynamics in glass-formers

Robert L. Jack,<sup>1</sup> Andrew J. Dunleavy,<sup>2,3,4</sup> and C. Patrick Royall<sup>2,3,5</sup>

<sup>1</sup>*Department of Physics, University of Bath, Bath, BA2 7AY, United Kingdom*

<sup>2</sup>*HH Wills Physics Laboratory, Tyndall Avenue, Bristol, BS8 1TL, UK*

<sup>3</sup>*School of Chemistry, University of Bristol, Cantock Close, Bristol, BS8 1TS, UK*

<sup>4</sup>*Bristol Centre for Complexity Sciences, Bristol BS8 1TW, UK*

<sup>5</sup>*Centre for Nanoscience and Quantum Information, Tyndall Avenue, Bristol, BS8 1FD, UK*

We analyse the connections between structure and dynamics in two model glass-formers, using the mutual information between an initial configuration and the ensuing dynamics to compare the predictive value of different structural observables. We consider the predictive power of normal modes, locally favoured structures, and coarse-grained measurements of local energy and density. The mutual information allows the influence of the liquid structure on the dynamics to be analysed quantitatively as a function of time, showing that normal modes give the most useful predictions on short time scales while local energy and density are most strongly predictive at long times.

PACS numbers: 64.70.Q-, 05.40.-a

As supercooled liquids approach their glass transitions, structural relaxation slows down dramatically, but molecular configurations remain disordered and apparently random [1, 2]. However, computer simulations [3–8] and experiments [9, 10] show that liquid structure and dynamical relaxation are correlated in these systems, as predicted (or assumed) in several theories [11–17]. However, correlations between structure and dynamics do not by themselves imply a causal relationship [18]: other theories [19] assume that local structure plays only a peripheral role in dynamical relaxation. Correlations between structure and dynamics can be demonstrated at a microscopic level [3–8], by exploiting the dynamically heterogeneous nature of glassy relaxation [20]. That is, individual particles have different propensities for motion [3], depending on local structure. Here, we use *information theory* [21] to analyze the strength of these correlations, by measuring the extent to which structural measurements can be used to *predict* particle dynamics at subsequent times. This quantitative analysis provides a stringent test of proposed causal links between structural features and slow dynamics, in contrast to previous analyses based on restricted subsets of particles or snapshots of the system. In two model glass-formers, we find that coarse-grained measurements of energy and density [22–24] give the most predictive information for long times. In one of the models, we also find that vibrational modes [4–6, 16] are strongly correlated with motion on relatively short time scales. Compared to these effects, the correlation between dynamics and low energy (or low enthalpy) local structures is relatively weak.

We present results for the Kob-Andersen (KA) mixture of Lennard-Jones particles [25], and an equimolar five-component hard sphere (HS) mixture, which mimics colloidal suspensions [26]. Both systems contain particles of different sizes, with the diameter of the largest particles being  $\sigma = 1$  (which sets the unit of length). The KA

system evolves with overdamped (Monte Carlo) dynamics as in [27]; we focus on a temperature  $T = 0.5$ . The HS system evolves by event-driven molecular dynamics [28]; we consider volume fractions  $\phi$  in the range  $0.52 - 0.58$ . In both systems, we use  $\Delta t$  to indicate the fundamental unit of time. The relaxation at the state points that we consider is up to 3 decades slower than relaxation at the onset of glassy dynamics, where it is of order  $\Delta t$  (in both systems). Further system details are given in the Supporting Information (SI) [36].

To characterize particle dynamics in these systems, we define the dynamical propensity [3] of particle  $i$  as  $\mu_{i,t} = \langle |\mathbf{r}_i(t) - \mathbf{r}_i(t_0)|^2 \rangle_{\text{iso}}$ , where  $\mathbf{r}_i(t)$  is the particle position at time  $t$ , and the isoconfigurational average is calculated over many independent dynamical simulations, all with the same initial particle positions but with independent random initial velocities (and independent stochastic dynamics in the KA system). The role of the “lag time”  $t_0$  is discussed in SI [36]: we take  $t_0 \approx 0.1\Delta t$ . We use  $s_i$  to denote a structural measurement at time  $t = 0$ , which depends in general on particle  $i$  and all particles in its vicinity. To quantify the strength of the correlation between  $s_i$  and the dynamical propensity  $\mu_{i,t}$ , we use mutual information (MI) measurements [21]. The MI is defined as

$$I_t(\mu; s) = \sum_s \int d\mu p_t(\mu, s) \log_2 \frac{p_t(\mu, s)}{p_t(\mu)p(s)}, \quad (1)$$

where  $p_t(\mu, s)$  is the joint probability distribution of  $\mu$  and  $s$ , while  $p_t(\mu)$  and  $p(s)$  are its marginal distributions. We assumed here that  $s_i$  takes discrete values: for continuous attributes  $s_i$ , the sum over  $s$  replaced by an integral.

The MI gives “the average amount of information about the propensity  $\mu_{it}$  that is provided by a measurement of  $s_i$ ”. Since  $s_i$  depends only on the initial condition, the MI measures *predictive* information. The MI

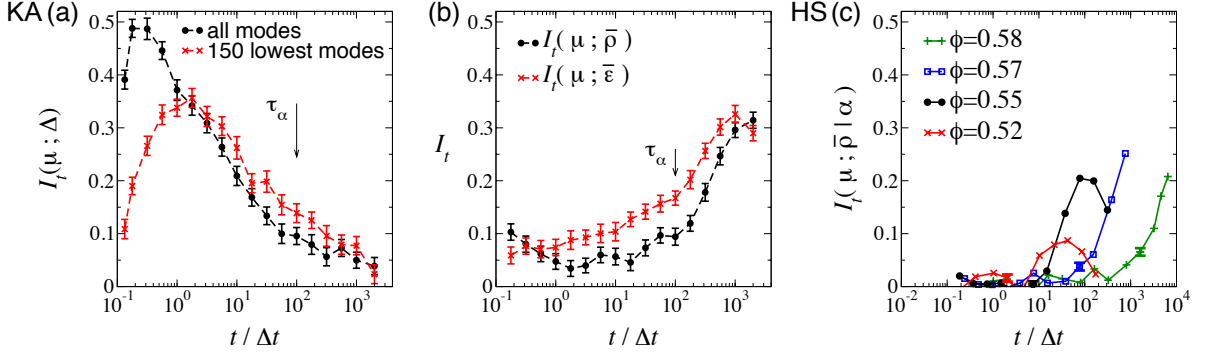


FIG. 1: MI measurements  $I_t(\mu; s)$  in the KA system and  $I_t(\mu; s|\alpha)$  in the HS system. We show MI between propensity and (a) Debye-Waller factors  $\Delta_i$  and  $\Delta_{i,n=150}$  in the KA system; (b) coarse-grained energy and density in the KA system; (c) coarse-grained density in the HS system. The structural relaxation time  $\tau_\alpha$  is indicated with arrows in (a,b). In (c), we show error bars only for  $t = \tau_\alpha$ : errors at other times are similar in magnitude. The behaviour of the MI at long times is discussed in the main text.

may be evaluated for any structural observable  $s_i$ , and it makes no assumptions on the nature of the correlation between  $\mu_{i,t}$  and  $s_i$ . As such, it represents a generally-applicable figure-of-merit for comparing the influence on dynamics of different structural measures, going beyond previous comparisons of snapshots [3–6, 22, 24] or analyses of selected subsets of particles [7, 8]. This use of (1) as a quantitative measure of information [21] is similar to the use of entropy as a measure of disorder in statistical mechanics, with the role of disorder being taken by the variation in propensity between different particles. Particles with the same value of  $s_i$  typically have less variation in their propensity, so specifying  $s_i$  reduces the variation in  $\mu_{it}$ , just as introducing a constraint in statistical mechanics reduces the entropy [29]. The MI is equivalent to this entropy reduction. Information is conventionally measured in bits, with one bit corresponding to a reduction in entropy of  $k_B \ln 2$ . Our procedure for estimating MI is described in the SI: the method ensures as far as possible that we obtain  $I_t(\mu; s) = 0$  if  $\mu$  and  $s$  are independent; it also provides an estimate of the numerical uncertainty in the MI.

To illustrate the use of MI, let  $s_i$  be the type (A or B) of particle  $i$  in the KA system. The different types have different dynamical relaxation so measuring the particle type provides predictive information about particle dynamics. In SI [36], we show that measuring the type of particle  $i$  provides between 0.1 and 0.7 bits of information about the propensity  $\mu_{i,t}$ , depending on the time  $t$ . This value is a useful baseline in interpreting the results that follow: if a structural measurement is strongly coupled with dynamics, we argue that  $I_t(\mu; s)$  should be at least of order 0.1 bit, while MIs much less than this are indicative of weak coupling.

Figure 1 shows MI measurements between particle propensities and several aspects of liquid structure, for both KA and HS systems. Since the influence of par-

ticle type on dynamics is not directly related to glassy behaviour, we measure mutual information where the predictability based on particle type has already been taken into account. That is, for several different  $s_i$ , we measure “the information about  $\mu_{i,t}$  that is provided by a measurement of  $s_i$ , for a particle whose type is already known”. In the KA system, we achieve this by restricting the distributions in (1) to particles of type A, which form the majority (80%) of the system. In the HS system, we use a ‘conditional MI’,  $I(\mu; s|\alpha) = \sum_{s,\alpha} \int d\mu p(\mu, s, \alpha) \log_2 \frac{p(\mu, s|\alpha)}{p(\mu|\alpha)p(s|\alpha)}$  where  $\alpha$  indicates the particle type [36]. Our choices of  $s_i$  reflect different theoretical pictures of glassy systems: we now discuss the implications of these results for those theories.

Several links have been proposed between normal modes in glassy systems and their material properties [4–6, 16]. Low-frequency modes in a supercooled liquid define a set of “soft directions” on its potential energy surface (or energy landscape), and both thermal fluctuations and structural relaxation couple significantly to these modes [4–6]. These modes also play a central role in the analogy between glassy behaviour and jamming [16]. We analyze them [36] by quenching the KA system to its nearest energy minimum (inherent structure), and diagonalizing the Hessian matrix of the energy at that minimum. The resulting eigenvectors and eigenvalues are  $\vec{v}_k$  and  $\omega_k^2$ , for  $k = 1 \dots 3N$ , and one defines a “local Debye-Waller (DW) factor”  $\Delta_i^2 = \sum_k |\mathbf{v}_k^i|^2 / \omega_k^2$  that indicates [6, 30] the expected size of fluctuations in the position of particle  $i$ , based on an expansion about the energy minimum. (Here  $\mathbf{v}_k^i$  is a vector containing the three components of  $\vec{v}_k$  associated with particle  $i$ .) Since low frequency modes couple most strongly to structural relaxation [6], we also define a generalised DW factor  $\Delta_{i,n}^2$ , which is calculated using only the  $n$  modes with lowest  $\omega_k$ . In HS systems, normal modes cannot be defined by reference to a potential energy surface so we do

not consider them here, although alternative definitions of normal modes are possible [5, 16].

Figure 1(a) shows that for relatively short time scales  $t \approx \Delta t \ll \tau_\alpha$  in the KA model, the mutual information between propensity and DW factors is large (up to 0.5 bits), so  $\Delta_i^2$  and  $\Delta_{i,n=150}^2$  are strongly correlated with particle motion. This indicates that the normal modes accurately mimic the fluctuations of the system within its initial metastable state. On longer time scales, the information provided by these measurements decreases strongly, but  $\Delta_{i,n=150}^2$  still provides more than 0.1 bits at the structural relaxation time  $\tau_\alpha$ , confirming that the low frequency normal modes do have significant predictive power for structural relaxation [5, 6, 16].

Coarse-grained energy and density measurements are also correlated with dynamical fluctuations [22–24, 31]. We define a local density, coarse-grained on a scale  $\ell$ , as  $\bar{\rho}_i^\ell = \ell^{-3} \sum_j e^{-r_{ij}^2/\ell^2}$ , where the sum runs over all particles  $j$  and  $r_{ij}$  is the distance between particles  $i$  and  $j$  [24]. Similarly, the locally-averaged energy is  $\bar{\varepsilon}_i^\ell = \sum_j \varepsilon_j e^{-r_{ij}^2/\ell^2} / (\ell^3 \bar{\rho}_i^\ell)$  where  $\varepsilon_j$  is the energy of particle  $j$ . Figures 1(b,c) show that for  $\ell = 2$  these coarse-grained quantities have strong predictive power on time scales longer than the structural relaxation time, but the MI is smaller for relaxation times up to and including  $\tau_\alpha$ . The results are broadly similar for both models (for the HS model, error bars are shown only at  $t = \tau_\alpha$ , to indicate the  $\phi$ -dependence of this time scale). We show data for  $\ell = 2\sigma$  since this gives a significant MI throughout this range of data: dependence of the MI on  $\ell$  is discussed in SI [36].

Throughout the glassy regime, we expect  $I(\mu; \bar{\rho})$  and  $I(\mu; \bar{\varepsilon})$  to have peaks at some time  $t^*$ , before decreasing at longer times (see for example the HS data at  $\phi = 0.55$ ). However, for the largest volume fractions it is clear that  $t^*$  is significantly larger than  $\tau_\alpha$ , and is larger than our sampling window. We attribute this large  $t^*$  to hydrodynamic effects that are largely independent of glassy behavior: regions of size  $\ell$  with high density or low energy relax on a time scale  $\ell^2/D$  where  $D$  is a diffusion constant. One therefore expects relaxation in such regions to be predictably slower than average up to times  $t^* \approx \ell^2/D$ , which is significantly larger than  $\tau_\alpha$ . Our focus here is on predictability on time scales of order  $\tau_\alpha$ , where the system is has significant dynamical heterogeneity and the motion is complex and co-operative. For this reason, we have not explored the large-time hydrodynamic behaviour in detail. We do note that for the HS system, the MI at  $\tau_\alpha$  increases at large  $\phi$ , indicating that the coupling of dynamics to local density is increasing as the glass transition is approached, consistent with [17, 23, 31]. However, even for the largest  $\phi$ , the MI is less than 0.1 bit at  $\tau_\alpha$ , although it does grow rapidly for larger times.

An alternative picture of glassy relaxation is based

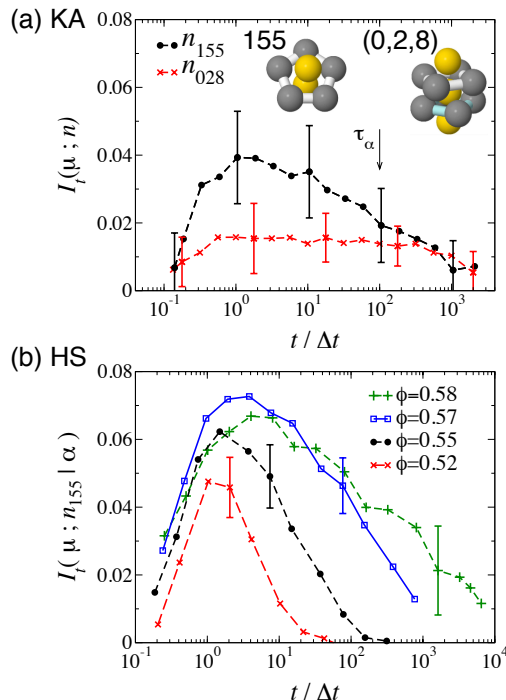


FIG. 2: MI between propensity and LFS measurements. We emphasise that these MIs are smaller in magnitude than those of Fig. 1, and all are less than 0.1 bit. (a) KA system, using LFS with ‘155’ and (0,2,8) signatures, as described in the main text and illustrated in the figure. (b) HS system, using LFS with CNA-155 signatures. Representative error bars are shown in (a), while in (b) we show error bars only at  $t = \tau_\alpha$ .

around locally-favoured structures (LFS): atomic or molecular packings that have low energy (or enthalpy) [14]. Particles in LFS typically have slower than average dynamics in glassy systems [7, 8, 32, 33]. For both KA and HS models, we consider an LFS based on a pentagonal bipyramid, identified by the signature 1551 in the analysis of [34]. These structures are indicative of local fivefold symmetry [35]. Let  $n_{155}(i)$  be the number of pentagonal bipyramids in which particle  $i$  participates [36]: we expect larger  $n_{155}(i)$  to be associated with lower propensity for motion. In the KA model, we also consider an LFS (bicapped square antiprism) that has been found to be correlated with slow dynamics [7, 8]. These LFS are associated with Voronoi polyhedra whose signature is (0,2,8) in the notation of [7]. We define  $n_{028}(i) = 1$  if particle  $i$  participates in such an LFS, with  $n_{028}(i) = 0$  otherwise. For the KA model, we calculate  $n_{155}$  and  $n_{028}$  using the inherent structure of the system.

Figure 2(a) shows results for the KA model, indicating that  $n_{155}$  and  $n_{028}$  are correlated with particle motion [7, 8]. As with the low-frequency normal modes, the signal is largest on time scales  $t \approx \Delta t$  indicative of  $\beta$ -relaxation, but there is still some correlation at the structural relaxation time. However, the strength of the

correlation is smaller for the LFS than for the normal modes, less than 0.1 bit in all cases. Figure 2(b) shows similar results for the HS system. The MI values are larger than those of the KA system, indicating that LFS have more predictive power for dynamics. At short times, the MI increases with increasing volume fraction; however the MI at  $\tau_\alpha$  (indicated by the error bars) depends more weakly on  $\phi$ . We argue that the small MI values at  $\tau_\alpha$  and longer times, and the absence of an increase of the values at  $\tau_\alpha$  with volume fraction, both indicate that the LFS identified here are more weakly coupled to the dynamics than the normal modes, at least for the degree of supercooling accessed here.

To summarize our findings so far, Figure 1 shows that Debye-Waller factors and coarse-grained measurements of energy and density have significant coupling to dynamics, providing predictive information comparable with measurements of particle type in the KA model. However, the information available from the different measurements has very different time-dependences. For short times, the normal mode analysis captures fast vibrational motion accurately, but the predictive power of this analysis decreases strongly with time. This indicates that as structural relaxation starts to take place, the ‘soft directions’ for further relaxation quickly diverge from those that were present at  $t = 0$ . On the other hand, coarse-grained energy and density measurements have almost no predictive value at short times, but the slow decay of large-scale hydrodynamic fluctuations means that they can influence particle dynamics quite strongly even on time scales much longer than  $\tau_\alpha$ : on these time scales, almost all memory of the initial structure has been lost, leaving only the hydrodynamic fluctuations in energy or density. For the state points considered here, Figure 2 shows that LFS measurements have less predictive power for dynamics in these models, and that this predictive power is largest on relatively short time scales associated with  $\beta$ -relaxation. Our interpretation is that since the lifetimes of most LFS are less than  $\tau_\alpha$  [8, 33], the influence of LFS on dynamics is (in most cases) similarly short-lived, limiting the predictive power of such measurements for dynamics.

Finally, in contrast to the measurements so far, we show how information theory can also be used to analyze how predictable particle motion is in these models, independent of any specific structural observable. Let  $p_{i,t}(r)$  be the (isoconfigurational) distribution of particle displacements  $r_i = |\mathbf{r}_i(t) - \mathbf{r}_i(t_0)|$ . Given data for  $N_p$  particles (which may be obtained in general from many initial configurations), we define

$$I_t(r; \text{id}) = N_p^{-1} \sum_i \int dr p_{i,t}(r) \log \frac{p_{i,t}(r)}{N_p^{-1} \sum_i p_{i,t}(r)}, \quad (2)$$

which is the “average amount of information about a particle’s motion that is provided by specifying its initial

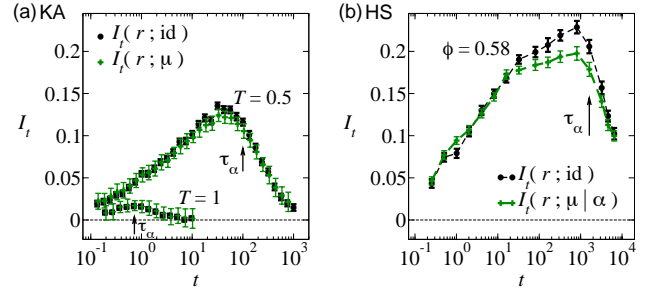


FIG. 3: Measurements of MI based on particle displacements  $r_i$ . (a) KA system for  $T = 0.5$  and  $T = 1$ ; (b) HS system at  $\phi = 0.58$ . See text for discussion.

environment”. Since a particle’s initial environment encodes all predictable aspects of its future motion,  $I_t(r; \text{id})$  indicates how predictable (or reproducible) particle motion is within the system [3, 24]. Figure 3 shows that  $I_t(r; \text{id})$  is much larger at low temperatures in the KA model than at high temperatures, indicating that structure is more strongly coupled to dynamics at low temperatures.

It is useful to compare  $I_t(r; \text{id})$  with “the average amount of information about a particle’s dynamics that is provided by specifying its propensity”, which is  $I_t(r; \mu) = \int dr d\mu p_t(r, \mu) \log \frac{p_t(r, \mu)}{p_t(\mu)p_t(r)}$ , where  $p_t(r, \mu)$  is the joint distribution of displacement  $r$  and propensity  $\mu$ , and  $p_t(r)$  is the marginal distribution of the displacement. Since fixing a particle’s initial environment necessarily fixes its propensity, one has

$$I_t(r; \mu) \leq I_t(r; \text{id}). \quad (3)$$

From Fig. 3, the two quantities in (3) are almost equal for the KA model. Eq. (3) is an “information-processing inequality” [21], so this result indicates that the propensity captures almost all predictable information about single-particle displacements. For the HS system, we use a conditional MI between  $r$  and  $\mu$ , to account for particle type, as above. The two MIs in (3) differ somewhat more strongly than they do in the KA model: this situation might arise (for example) if some particles have finite average displacements  $\langle \mathbf{r}_i(t) - \mathbf{r}_i(t_0) \rangle$  that are only weakly correlated with their propensities.

Nevertheless, we have  $I_t(r; \mu) \approx I_t(r; \text{id})$  in both models, indicating that the propensity captures all predictable (reproducible) aspects of the single-particle dynamics [3]. This further validates the use of the mutual information as a general figure-of-merit for evaluating proposed connections between structure and dynamics. Given the implications of Figs. 1 and 2 for the strength and time-dependence of the coupling between structure and dynamics, we hope that future studies will exploit these information-theoretic measurements to further elucidate which (if any) structural features are responsible for the strong dynamical slowing in supercooled liquids.

We thank Peter Harrowell, Peter Sollich, Gilles Tarjus, and Karoline Wiesner for helpful discussions. RLJ and AJD were supported by the EPSRC through grants EP/I003797/1 and EP/E501214/1 respectively. CPR gratefully acknowledges the Royal Society for financial support.

- 
- [1] M. D. Ediger, C. A. Angell and S. R. Nagel, *J. Phys. Chem.* **100**, 13200 (1996).
  - [2] P. G. Debenedetti and F. H. Stillinger, *Nature* **410**, 259 (2001).
  - [3] A. Widmer-Cooper, P. Harrowell and H. Fynewever, *Phys. Rev. Lett.* **93** 135701 (2004); A. Widmer-Cooper and P. Harrowell, *J. Chem. Phys.* **126** 154503 (2007).
  - [4] D. Coslovich and G. Pastore, *Europhys. Lett.* **75**, 7840 (2006).
  - [5] C. Brito and M. Wyart, *J. Stat. Mech.* (2007) L08003.
  - [6] A. Widmer-Cooper, H. Perry, P. Harrowell and D. R. Reichman, *Nature Physics* **4**, 711 (2008).
  - [7] D. Coslovich and G. Pastore, *J. Chem. Phys.* **127**, 124504 (2007).
  - [8] A. Malins, J. Eggers, H. Tanaka and C. P. Royall, *Faraday Discuss.* **167**, in press (2013).
  - [9] M. Leocmach and H. Tanaka, *Nature Communications* **3**, 974 (2012).
  - [10] R. Candelier, A. Widmer-Cooper, J. K. Kummerfeld, O. Dauchot, G. Biroli and D. R. Reichman, *Phys. Rev. Lett.* **105**, 135702 (2010).
  - [11] V. Lubchenko and P. G. Wolynes, *Ann. Rev. Phys. Chem.* **58**, 235 (2007).
  - [12] J.-P. Bouchaud and G. Biroli, *J. Chem. Phys.* **121**, 7347 (2004).
  - [13] W. Götze and L. Sjögren, *Rep. Prog. Phys.* **55**, 241 (1995).
  - [14] G. Tarjus, S. A. Kivelson, Z. Nussinov and P. Viot, *J. Phys.: Cond. Matt.* **17**, R1143 (2005).
  - [15] T. Kawasaki, T. Araki and H. Tanaka, *Phys. Rev. Lett.* **99**, 215701 (2007).
  - [16] M. L. Manning and A. J. Liu, *Phys. Rev. Lett.* **107**, 108302 (2011).
  - [17] G. Biroli, J.P. Bouchaud, K. Miyazaki, and D. R. Reichman, *Phys. Rev. Lett.* **97**, 195701 (2006).
  - [18] D. J. Ashton and J. P. Garrahan, *Eur J. Phys. E* **30**, 30 (2009).
  - [19] J. P. Garrahan and D. Chandler, *Proc. Nat. Acad. Sci. USA* **100**, 9710 (2003); D. Chandler and J. P. Garrahan, *Ann. Rev. Phys. Chem.* **61**, 191 (2010).
  - [20] M. D. Ediger, *Annu. Rev. Phys. Chem.* **51**, 99 (2000).
  - [21] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
  - [22] G. S. Matharoo, M. S. G. Razul, and P. H. Poole, *Phys. Rev. E* **74**, 050502 (2006).
  - [23] L. Berthier, G. Biroli, J.-P. Bouchaud, W. Kob, K. Miyazaki and D. R. Reichman, *J. Chem. Phys.* **126** 184503 (2007).
  - [24] L. Berthier and R. L. Jack, *Phys. Rev. E* **76**, 041509 (2007).
  - [25] W. Kob and H. C. Andersen, *Phys. Rev. E* **51**, 4626 (1995); **52**, 4134 (1995).
  - [26] C. P. Royall, A. Malins, A. J. Dunleavy, and R. Pinney, in *Fragility of glass-forming liquids*, editors: K. A. Kelton, L. Grier, S. Sastry (Hindustan book agency, New Delhi, 2014).
  - [27] L. Berthier and W. Kob, *J. Phys: Cond. Matt* **19**, 205130 (2007).
  - [28] M. N. Bannerman, R. Sargant and L. Lue, *J. Comp. Chem.* **32**, 3329 (2011).
  - [29] D. Chandler, *An introduction to modern statistical mechanics*, (OUP, Oxford, 1987).
  - [30] A. Widmer-Cooper and P. Harrowell, *Phys. Rev. Lett.* **96**, 185701 (2006).
  - [31] L. Berthier, G. Biroli, J.-P. Bouchaud, L. Cipelletti, D. El Masri, D. L'Hôte, F. Ladieu and M. Pierno, *Science* **310**, 1797 (2005).
  - [32] C. P. Royall, S. R. Williams, T. Ohtsuka and H. Tanaka, *Nature Materials* **7**, 556 (2008).
  - [33] A. Malins, J. Eggers, C. P. Royall, S. R. Williams and H. Tanaka, *J. Chem. Phys.* **138**, 12A535 (2013).
  - [34] J. D. Honeycutt and H. C. Andersen, *J. Phys. Chem* **91**, 4950 (1987).
  - [35] H. Jonsson and H. C. Andersen, *Phys. Rev. Lett.* **60**, 2295 (1988).
  - [36] See Supporting Information.

## Supporting Information

This supporting information contains:

- Details of the models described in the main text, and the methods used to identify locally-favored structures.
- Illustrative results of mutual information between propensity and particle type
- Discussion of the  $\ell$ -dependence of the results shown in Fig. 1(b,c)
- Analysis of the numerical method that we use when estimating mutual information.

### Model systems

The KA mixture is defined as in [25]. The system consists of  $N = 1400$  particles of which 80% are of type A and 20% of type B. The particles interact by Lennard-Jones potentials with parameters  $(\epsilon_{AA}, \epsilon_{AB}, \epsilon_{BB}) = (1.0, 1.5, 0.5)\epsilon$  and  $(\sigma_{AA}, \sigma_{AB}, \sigma_{BB}) = (1.0, 0.8, 0.88)\sigma$ . Temperatures are quoted in units of  $\epsilon$ , with Boltzmann's constant  $k_B = 1$ . The onset temperature of glassy dynamics is  $T \approx 1$  and the mode-coupling temperature has been estimated [25] to be  $T = 0.435$ . The total number density of particles is  $(10/9.4\sigma)^3 \approx 1.2\sigma^{-3}$ , and we use a cubic simulation box with periodic boundaries. The system evolves by Monte Carlo dynamics as in [27], with trial displacements drawn from a cube of side  $0.15\sigma$  centred at the origin. The mean-square displacement per trial move is  $\delta^2 = (0.075\sigma)^2$  and we define the fundamental time unit  $\Delta t = \sigma^2/D_0$  where  $D_0$  is the diffusion constant of free particle. The result is that  $\Delta t$  corresponds to  $6(\sigma/\delta)^2 \approx 1070$  proposed MC moves per particle. The structural relaxation time  $\tau_\alpha$  discussed in the main text is defined as  $F_s(k, \tau_\alpha) = (1/e)$  where  $F_s(k, t) = \langle e^{-i\mathbf{k} \cdot [\mathbf{r}_i(t) - \mathbf{r}_i(0)]} \rangle$  is evaluated by an average over particles of type A, and  $k = |\mathbf{k}| = 7.25/\sigma$ .

The polydisperse hard sphere system consists of an equimolar mix of five particle species with diameters  $(1.000, 0.938, 0.899, 0.861, 0.799)\sigma$ , all with equal masses  $m$ . The particles interact as hard spheres and the system evolves by event-driven molecular dynamics (implemented by DynamO [28]). The system comprises  $N = 1372$  particles and the simulation box is cubic with periodic boundary conditions. The time unit in the system is  $\Delta t = \sqrt{m\sigma^2/k_B T}$ . The structural relaxation time is evaluated at  $k = 2\pi/\sigma$ .

### Identifying locally favored structures

Here, we briefly describe the structural measurements  $n_{155}$  and  $n_{028}$  that we use to identify locally-favoured

structures in these systems. These measurements are based on Voronoi analyses of the system. In the KA model, we perform this analysis after quenching the system to its nearest energy minimum (inherent structure). We follow [7] in using a Voronoi analysis where faces between A and B particles are located closer to the B particles, consistent with their smaller size. In the HS system, we use a regular Voronoi analysis, in which faces are midway between neighbouring particles.

We identify  $(0, 2, 8)$  Voronoi polyhedra in the KA system as those with ten faces, of which exactly two have four edges, and eight have five edges. This particle and its ten Voronoi neighbours form a cluster (11A in the topological cluster classification [S1]), and we set  $n_{028}(i) = 1$  for all particles in these clusters.

To identify the pentagonal bipyramids in which particle  $i$  participates (in both HS and KA systems), we identify  $n_{155}(i)$  as the number of pentagonal faces on the Voronoi cell of that particle. This gives the number of neighbours of particle  $i$  that share exactly five mutual neighbours with particle  $i$ . The procedure is equivalent to counting the number of '1551' bonds in the common neighbour analysis (CNA) [34,35], and is similar to the identification of '7A' clusters in the topological cluster classification [S1].

### MI between particle type and propensity

To demonstrate the physical meaning of MI, Fig. S1(a) shows  $I_t(\mu; s)$  for the KA system, where the structural measurement  $s_i$  is taken to be the particle type  $\alpha_i = A, B$ . The B-particles are more mobile in this system, and as shown in the inset, at large times  $t \gg \tau_\alpha$ , the propensity distributions for the two kinds of particle have almost zero overlap. Thus, for these very long times, measuring the particle type splits the propensity distribution into two distinct components: this provides  $-f \log_2 f - (1-f) \log_2 (1-f)$  bits of information (similar to a mixing entropy), where  $f$  and  $(1-f)$  are the fractions of particles in each component. If the components were equal in size, the MI would be exactly 1 bit: here the B particles are less numerous ( $f = 0.2$ ) so the MI is less, approximately 0.7 bits. For times  $t$  close to the structural relaxation time  $\tau_\alpha$ , Fig. S1(b) shows that the propensity distributions of the two types differ from each other, but there is a region of significant overlap. In this case, measuring the particle type provides 0.3 bits of information about  $\mu_{it}$ .

### Predictive power of local energy/density: dependence on length scale $\ell$

The coarse-grained measurements of local energy and density  $\bar{\rho}$  and  $\bar{\epsilon}$  discussed in the main text depend (by



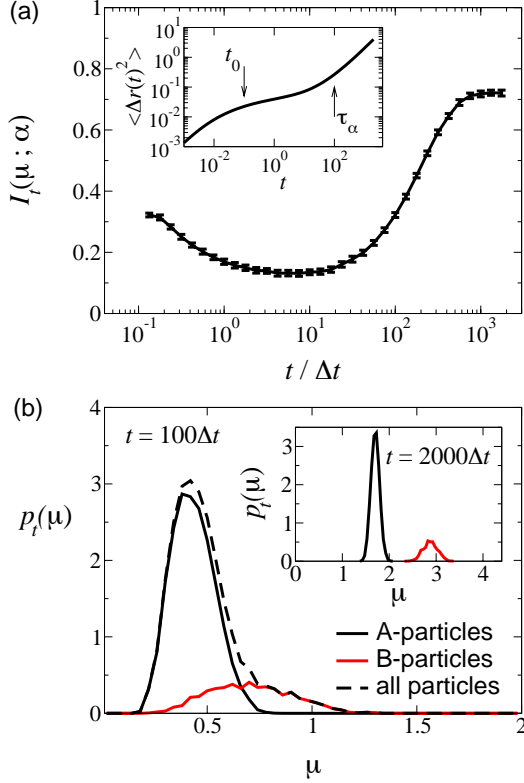


FIG. S1: Illustration of MI measurements. (a) Mutual information between the propensity  $\mu$  and the particle type  $\alpha$  in the KA system. (Inset) Mean square displacement of all particles, with arrows indicating the structural relaxation time  $\tau_\alpha \approx 100\Delta t$  and the lag time  $t_0 = 0.1\Delta t$ . (b) Distributions of the propensity at  $t = 100\Delta t$  and  $t = 2000\Delta t$  (inset): see text for discussion.

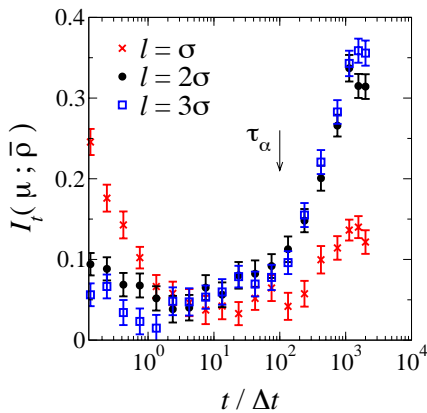


FIG. S2: MI between coarse-grained density  $\bar{\rho}^\ell$  and propensity in the KA model at  $T = 0.5$ , for  $\ell$  between  $\sigma$  and  $3\sigma$ . [The data for  $\ell = 2\sigma$  are also shown in Fig. 2(c) of the main text.] See text for discussion.

definition) on a length scale  $\ell$  that indicates the size of the local coarse-graining region. The effect of varying this length scale is illustrated in Figure S2, where the MI between propensity and  $\bar{\rho}^\ell$  is plotted, for various  $\ell$ . For short lengths (here,  $\ell = \sigma$ ), the local density is strongly predictive on short times, presumably due to the influence of free volume on the vibrational motion of particles within a single metastable state. For longer times, the MI increases with  $\ell$  before saturating at  $\ell \approx 2\sigma$ . In the long time regime, and for all cases considered, the MI for  $\ell = 2\sigma$  is almost always larger than for smaller  $\ell$ -values, and increasing  $\ell$  above  $2\sigma$  does not significantly increase the MI (as in Fig. S2). This observation motivated our choice of  $\ell = 2\sigma$  for the data shown in Figs. 2(c,f). The MI between coarse-grained energy and propensity does not show the early-time signal found for  $I(\mu; \bar{\rho})$  at  $\ell = \sigma$  but otherwise behaves similarly to  $I(\mu; \bar{\rho})$ .

### Estimating mutual information

Calculating mutual information (MI) from numerical data requires some care, since estimators are vulnerable to systematic errors if sample sizes are not sufficiently large. A variety of estimators have been developed (see for example [S1-S6]), many of which use Bayesian methods, exploiting prior knowledge (or assumptions) about the form of the underlying distributions in order to better estimate either entropies or mutual informations [S4, S6, S7]. In this work, we use a simple method that we have tailored to the problem of interest here, based on the method of [S5].

In all measurements, we discretise the propensity, forming a histogram with bins of width  $\delta\mu = \langle\mu\rangle/10$ , where  $\langle\mu\rangle$  is the mean propensity. The width of the bins is comparable with the numerical uncertainties in our estimates of the  $\mu_i$ , which are obtained from between 100 and 250 independent trajectories. For this reason, storing the propensities to greater accuracy than the bin width would not make our measurements of MI any more accurate – the binning does not introduce numerical artefacts, and is convenient in what follows. Further, since the same binning is used for all MI measurements, we are able to make a fair comparison between the different structural measurements shown in Figures 1 and 2 of main text. In the following, we use  $m_i$  as an integer-valued label for the bin in which the propensity  $\mu_i$  is located (for example, one may take  $m_i = \lfloor \mu_i/\delta\mu \rfloor$ , the largest integer that is less than or equal to  $\mu_i/\delta\mu$ ).

### Two discrete variables

We first describe the estimator that we use for calculating MI between two discrete-valued variables. We have in mind that  $s_i$  is a structural observable with a discrete set

of possible values, while  $m_i$  is the propensity bin-index as described above. However, the discussion is general for joint distributions of discrete random variables. For each particle, suppose that we measure two integers  $m_i$  and  $s_i$ . Then given data for  $N_p$  particles, let  $n(m, s)$  be the number of particles with  $(m_i, s_i) = (m, s)$ ; also let  $n(m) = \sum_s n(m, s)$  be the number of particles with  $m_i = m$ , and similarly  $n(s) = \sum_m n(m, s)$ . The simplest MI estimate based on these data is the “plugin estimator”:

$$I^0 = N_p^{-1} \sum_{m,s} n(m, s) \log_2 \frac{n(m, s) N_p}{n(m) n(s)} \quad (4)$$

where the sum runs over all pairs  $(m, s)$  for which  $n(m, s) > 0$ . Given sufficient data,  $I^0$  converges to the mutual information  $I(m; s)$ , however, this convergence is often quite slow, requiring very large  $N_p$  for an accurate estimate. In particular, even if the data set is constructed so that  $m_i$  and  $s_i$  are independent, one typically finds  $I^0 > 0$ , recovering  $I^0 \rightarrow 0$  only as  $N_p \rightarrow \infty$ .

To see the reason for this, it is useful to write  $I^0 = H^0(m) - H^0(m|s)$  with

$$H^0(m|s) = - \sum_s \frac{n(s)}{N_p} \sum_m n(m|s) \log_2 n(m|s) \quad (5)$$

where  $n(m|s) = n(m, s)/n(s)$ , and

$$H^0(m) = - \sum_m \frac{n(m)}{N_p} \log_2 \frac{n(m)}{N_p} \quad (6)$$

The key point is that for large enough data sets ( $N_p \rightarrow \infty$ ) one has  $n(m)/N_p \rightarrow p(m)$  by the law of large numbers, so that  $H^0(m)$  is an entropy estimator for  $H(m) = - \sum_m p(m) \log_2 p(m)$ . Similarly,  $H^0(m|s)$  converges to a weighted sum of conditional entropies of the form  $\sum_m p(m|s) \log_2 p(m|s)$ , as long as  $n(s) \rightarrow \infty$  for all  $s$ . The difficulty is that the convergence of  $H^0(m)$  and  $H^0(m|s)$  to their respective limits are ruled by different large parameters ( $N_p$  and  $n(s)$ ), and there are systematic errors associated with this convergence if these parameters are not large enough. In general,  $H^0(m)$  and  $H^0(m|s)$  both underestimate the relevant entropies, but the error on  $H^0(m)$  is smaller, resulting in a positive systematic error for  $I^0(m; s)$ .

To reduce this effect, we define an alternative estimator

$$I^1 = \sum_s \frac{n(s)}{N_p} \sum_m n(m|s) \left[ \log_2 n(m|s) - \log_2 \frac{\tilde{n}_s(m)}{n(s)} \right] \quad (7)$$

Here, the  $\tilde{n}_s(m)$  are obtained from a random null data set, as follows. For each  $s$ , we draw a random sample of size  $n(s)$  (without replacement) from the original data set, and  $\tilde{n}_s(m)$  is defined as the number of particles in that random sample that have  $m_i = m$ . For a large enough sample,  $\tilde{n}_s(m)/n(s) \rightarrow n(m)/N_p \rightarrow p(m)$ .

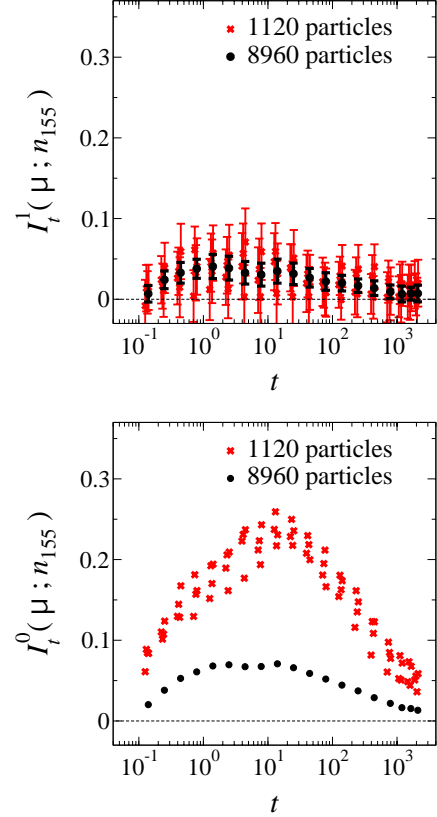


FIG. S3: (a) Estimate of MI between  $n_{155}$  and the propensity, using the estimator  $I^1$ . We show one estimate based on a sample of  $(8 \times 1120) = 8960$  type-A particles (from 8 independent initial conditions), and four estimates each based on 1120 particles (each estimate based on just one initial condition). The error bars are conservative estimates of numerical uncertainty (see text) and are reduced on including more data. The mean estimate has only a weak drift with sample size indicating that systematic errors arising from small samples are small. (b) Estimates of the same MI based on the same data but using the estimator  $I^0$ . The estimation is poor in this case: there is a strong systematic error coming from the finite sample size. The error decreases as more data is included but the largest sample used is not large enough to saturate the limit  $N_p \rightarrow \infty$  and does not provide a reliable estimate of the MI.

However, the advantage of the method is that the convergence of  $n(m|s)$  to  $p(m|s)$  and  $\tilde{n}_s(m)/n(s)$  to  $p(m)$  are now ruled by the same parameter  $n(s)$ , and the result is that the systematic errors arising from the two terms in (7) tend to cancel each other. We note that if instead of drawing a random sample we simply set  $\tilde{n}_s(m)/n(s) = n(m)/N_p$ , independent of  $s$ , then we recover the original estimator  $I^0$ . It is also notable that if  $m$  and  $s$  are independent then the conditional distribution  $n(m|s)$  should be statistically equivalent to the distribution obtained in the random sample  $\tilde{n}_s(m)$ . This means that  $I^1$  is free from systematic error in the case where  $s$  and  $m$  are independent. This is the most impor-



tant case for the calculations of this paper, because the MI values found are typically quite small, and systematic errors when  $I \approx 0$  correspond to false-positive signals of correlation between structure and dynamics, which can be misleading.

For each estimate of MI, we compute  $I^1$  using several random null data sets (typically 100 realisations are sufficient). The average value of  $I^1$  over the realisations provides our estimate of  $I$  while the standard deviation among the values of  $I^1$  gives an estimate on the uncertainty of this estimate. We therefore use this standard deviation as the error bar for the estimate of  $I$ . We emphasise that the  $n(m|s)$  are determined by the original data and are the same for every realisation of the null data: it is the finite size of this original data set that introduces a finite uncertainty on estimates of  $I$ . This uncertainty is not reduced by repeated sampling over different null data sets, so it is the standard deviation of  $I^1$  that gives the relevant error estimate, not the standard error.

Fig. S3 shows estimates for the MI between  $n_{155}$  and the propensity, obtained by the estimators  $I^1$  and  $I^0$ , for data sets of two different sizes. It can be seen that  $I^0$  is not sufficient for the purposes used here, even for the larger data set, while  $I^1$  gives a consistent estimate of the MI for data sets of both sizes considered. The estimate of the uncertainty based on  $I^1$  is also self-consistent, in that error bars from independent estimates typically overlap with each other.

An alternative to (7) can be obtained by interchanging  $s$  and  $m$ , since the MI is symmetric:

$$I'^1 = \sum_m \frac{n(m)}{N_p} \sum_s n(s|m) \left[ \log_2 n(s|m) - \log_2 \frac{\tilde{n}_m(s)}{n(m)} \right] \quad (8)$$

We use  $I^1$  for all calculations of MI between discrete variables, but it is useful to define  $I'^1$  in preparation for later sections.

#### One discrete and one continuous variable

We now turn to the case where the structural variable of interest takes continuous values. The analogue of the estimator  $I^1$  in (8) is

$$I^2 = \sum_m \frac{n(m)}{N_p} F_{s|m} \quad (9)$$

where  $F_{s|m}$  is an estimator for  $\int ds [p(s|m) \log_2 p(s|m) - p(s) \log_2 p(s)]$ . To define  $F_{s|m}$ , we again take a random sample of size  $n(m)$  from the original data, to ensure that systematic errors on estimates for  $\log p(s|m)$  and  $\log p(s)$  should cancel as far as possible. Then, if  $s^1, s^2, \dots$  is an ordered list of the values of  $s_i$  for those particles with  $m_i = m$ , and  $\tilde{s}^1, \tilde{s}^2, \dots$  is an ordered list of the values of

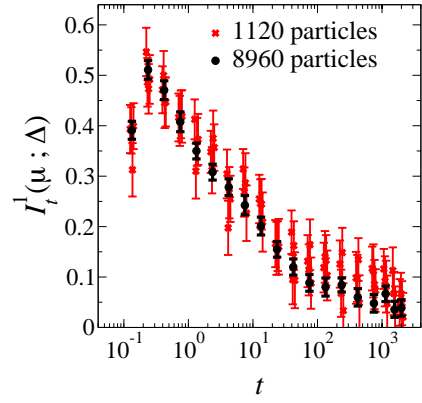


FIG. S4: Estimate of MI between  $\Delta_i^2$  and the propensity, using the estimator  $I^2$ . We show one estimate based on a sample of  $(8 \times 1120) = 8960$  type-A particles (from 8 independent initial conditions), and four estimates each based on 1120 particles (from independent initial conditions). The error bars are conservative estimates of numerical uncertainty (see text) and are reduced on including more data. Systematic errors from small samples are weak.

$s_i$  in the random sample, we define

$$F_{s|m} = \frac{1}{[n(m) - 1] \ln 2} \sum_{i=1}^{n(m)-1} [\ln(s^{i+1} - s^i) - \ln(\tilde{s}^{i+1} - \tilde{s}^i)]. \quad (10)$$

This converges to the required result as  $n \rightarrow \infty$  because if  $x^1, x^2, \dots, x^n$  is an ordered list of independent random samples from  $p(x)$  then, as  $n \rightarrow \infty$ , one has

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \ln(x^{i+1} - x^i) + \psi(1) - \psi(n) \rightarrow - \int dx p(x) \ln p(x) \quad (11)$$

where  $\psi(n)$  is the digamma function, which satisfies  $\psi(n+1) - \psi(n) = 1/n$  and  $\psi(1) = -\gamma$  where  $\gamma = 0.577 \dots$  is Euler's constant [S5].

Fig. S4 shows results using the estimator  $I^2$ . As with  $I^1$ , the uncertainty on the estimate of MI is reduced on including more data, and the systematic variation on increasing  $N_p$  is weak, indicating that the estimator is reliable.

- 
- [S1] A. Malins, S. R. Williams, J. Eggers, and C. P. Royall, *J. Chem. Phys.* **139**, 234506 (2013)..
- [S2] For a review, see K. Hlavackova-Schindler, M. Palus, M. Vejmelka and J. Battacharya, *Phys. Rep.* **441**, 1 (2007).
- [S3] T. Schürmann and P. Grassberger, *Chaos* **6**, 414 (1996)
- [S4] I. Nemenman, F. Shafee and W. Bialek, in *Advances in Neural Information Processing Systems* **14**, eds. T.G. Dietterich, S. Becker and Z. Ghahramani (MIT Press, Cambridge, 2002).

- [S5] A. Kraskov, H. Stögbauer and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).
- [S6] M. B. Kennel, J. Shlens, H. D. I. Arbanel and E. J. Chichilnisky, Neural Comp. **17**, 1531 (2005)
- [S7] E. Archer, I. M. Park, and J. W. Pillow, Entropy **15**, 1738 (2013)